# Learn to Explain: Multimodal Reasoning via Thought Chains for Science Question Answering

## Abstract

When answering a question, humans utilize the information available across different modalities to synthesize a consistent and complete chain of thought (CoT). This process is normally a black box in the case of deep learning models like large-scale language models. Recently, science question benchmarks have been used to diagnose the multi-hop reasoning ability and interpretability of an AI system. However, existing datasets fail to provide annotations for the answers, or are restricted to the textual-only modality, small scales, and limited domain diversity. To this end, we present Science Question Answering (ScienceQA), a new benchmark that consists of ~21k multimodal multiple choice questions with a diverse set of science topics and annotations of their answers with corresponding lectures and explanations. We further design language models to learn to generate lectures and explanations as the chain of thought (CoT) to mimic the multi-hop reasoning process when answering ScienceQA questions. ScienceQA demonstrates the utility of CoT in language models, as CoT improves the question answering performance by 1.20% in few-shot GPT-3 and 3.99% in fine-tuned UnifiedQA. We also explore the upper bound for models to leverage explanations by feeding those in the input; we observe that it improves the few-shot performance of GPT-3 by 18.96%. Our analysis further shows that language models, similar to humans, benefit from explanations to learn from fewer data and achieve the same performance with just 40% of the data. The data and code are available at https://scienceqa.github.io.